**TASK ORDER 68HERH20F0406 UNDER**
**CONTRACT EP-C-17-017**

# EXTERNAL PEER REVIEW OF EPA'S, *A PROOF-OF-CONCEPT CASE STUDY INTEGRATING PUBLICLY AVAILABLE INFORMATION TO SCREEN CANDIDATES FOR CHEMICAL PRIORITIZATION UNDER TSCA*

## FINAL PEER REVIEW REPORT
## February 3, 2021

*Submitted to:*
**U.S. Environmental Protection Agency**
Office of Research and Development
Center for Computational Toxicology and Exposure
Research Triangle Park, NC
Attn: Andrew Watkins
Watkins.Andrew@epa.gov

*Submitted by:*
**Eastern Research Group, Inc.**
110 Hartwell Avenue
Lexington, MA 02421

ERG
www.erg.com

# CONTENTS

**COMMENTS SUBMITTED BY**

**Tara S. Barton-Maclaren, Ph.D.**

Sr Research Manager, Emerging Approaches Unit
Health Canada
Healthy Environments and Consumer Safety Branch
Ontario, Canada

**External Peer Review of *A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA***

**COMMENTS TO REVIEW QUESTIONS:**

Thank you for the opportunity to review the EPA TSCA Proof-of-Concept Case study. This is a well written and clear document outlining the data sources, methods and outcomes of the approach relative to the chemical space selected to illustrate the utility and sensitivity of the tool. Introducing automated approaches to enhance the efficiency of chemical screening and prioritization is an area under exploration and development is risk assessment programs internationally. The thoughtful integration of data and technical advancements described and tested in this PICS approach will provide a useful foundation to support modernizing approaches to priority setting in programs external to the EPA.

For transparency, I would like to acknowledge that my review focused on the details related to the human health hazard and exposure domains. The ecological domains are outside the scope of my scientific expertise.

1. **OVERALL QUESTIONS**
   **Based on your knowledge and understanding of toxicology and/or exposure, chemistry, and risk assessment, please comment on the overall TSCA POC document.**

**1A. Does this document address the purpose and aims as laid out in the introduction?**

In the introduction, the EPA effectively delineates the key drivers behind the need for a more effective, rapid and consistent approach to screen chemicals for further evaluation – that being far too large an inventory to continue to manually collect, synthesize and review information that can be more efficiently processed using automated workflows. It is explicitly outlined that the approach is to include both traditional and NAMs in the scientific domains. The addition of the background section with relevant information set the stage well delineating the motivation and requirements of the presented approach.

The purpose and aims are clearly defined; that being to understand the information landscape of large inventory of chemicals, provide transparent and reproducible, as well as a flexible and sustainable process, increase efficiency and manage workload, and create a modular workflow that can be readily adapted. Each aim is thoroughly met as evidenced by the detailed descriptions of the data used, the considerations for interpretation, the development and application of the domain criteria as well as by the recognition of limitations and areas for further work. What the approach is not intended to do was also stated. This is important to acknowledge and adds clarity on purpose and scope. For consideration as an additional element of "not intended" I do wonder if it should be acknowledged that this approach does not include the comprehensive screening of scientific literature that may be available and used for subsequent risk assessments (i.e. PubMed, SciFinder, etc). The document indicates throughout that this approach integrates publically-available information, for transparency acknowledging the scope of information that is captured up front would also be useful as this will inherently limit the information availability metric of many chemicals that may have data outside the scope of the sources used.

**1B. Are the ideas presented throughout the document clear and presented in a logical manner?**

Generally, the document is clearly written, well organized and the style and format makes the document and technical information easy to follow. The overall approach is introduced to provide an overarching view and each of the domains are systematically described. The use of consistent tables and the inclusion of definitions and brief explanations in footnotes adds further clarity and transparency in the document. The use of schematics and graphical representations of the information and results effectively communicates the outcomes of the proof-of-concept evaluation.

**1C. Is the method described in this document appropriate to be scalable to the thousands of chemicals on the TSCA inventory?**

Yes, given that the domain workflows are programmed to run in an automated fashion and search large publically-available curated datasets, I see no concerns in scaling this approach to address the TSCA inventory. Technology and the workflow, in my opinion, is not the limitation of the approach; this is what in fact makes the method very scalable to thousands of chemicals. Data availability from the Type 1 information sources used more broadly for the TSCA inventory is the more challenging element. In scaling the approach across the larger chemical space more emphasis may be needed on NAM data than was used in the proof of concept which included chemicals either with fairly high levels of information or those for which there is already an established level of concern.

**1D. Is this approach adaptable to other large-scale chemical prioritization efforts other than for TSCA?**

Yes, many other chemical risk assessment programs internationally are seeking methods to introduce efficiencies, transparency and consistency in their screening, prioritization and assessment processes. The approach described provides a first-tier screening approach for prioritization but may also find utility for other programs, with some modification, as a strategy for automated data searching and collection of data for those chemicals already designated for risk assessment. The design and inclusion of the information gathering flags nicely supports early identification of data gaps which can be leveraged early in the problem formulation or risk assessment process to direct further information gathering or data generation as appropriate. The EPA has gone to great efforts to compile and curate a significant amount of data that is and undoubtedly will continue to be useful for other regulatory programs. The fact that the approach is designed in domains and the data sources are well documented allows the flexibility for others to adapt and customize as needed to meet program requirements or needs that might be different to those of the EPA.

**2. SCIENTIFIC DOMAINS**

**Based on your knowledge of toxicology, chemistry, risk assessment, and/or exposure science, please comment on the evaluation, workflow, and metrics developed for the individual scientific domains in the TSCA POC.**

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

The domain specific evaluations were logical and weighted appropriately based on tiers of information with *in vivo* scoring highest. This aligns with current risk assessment requirements and the use of animal studies for hazard characterization and point of departure derivation.

Designing the domain workflows to select more conservative options is appropriate as a first tier in the screening and prioritization process. Further refinement can be done during the expert review stage. The specific identification of the study types that are missing with the IG flag is a key element of the

reporting in particular if there are opportunities to address the data gaps in advance of starting the risk assessment.

For the Hazard-to-Exposure (HER) evaluation, there would be added value to include a more detailed description related to the interpretation and impact of this metric in the overall outcome of the scientific domain results. Given that this is the metric that provides the risk-based context, I wondered if the weight of this metric in the overall scheme should have a more prominent role. I've included more specific comments related to the chemical comparison in the results section below.

A general observation regarding the scientific domains is that this is a sound data driven approach for substances that have a more fulsome dataset, existing hazard classifications and/or existing assessments. However, it is not as clear if the approach will support the identification of substances that have the potential to be of concern but that are lacking traditional (animal) data from the data sources currently incorporated. The goal of prioritizing substances with higher data availability has been effectively achieved. If there is a desire to also document those substances that (may) require further action but that do not have traditional data, then this may not have been achieved based on the designation of a value of 0 or IG flag when other than primary source animal data are considered. The noted exception is for the genotoxicity domain. Perhaps related to the comment above, this is where further details on the application of the HER, BER, TER could be expanded to better illustrate how this subset of chemicals will be defined. Expanding the POC test set to include chemicals based on the lower tiered data might help to delineate this aspect of the approach further.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

One element of the scientific domain that is not included is that of endocrine activity. The EPA has made much progress in this area in terms of developing tiered screening approaches for the prioritization of chemicals for potential endocrine disruption. The absence of this domain from the approach appears to be a gap that might also contribute to the expansion of considerations in the susceptible populations domain.

**Carcinogenicity Domain** - Based on the IG flags description in table 2 my interpretation is that there are a number of limited data sources that in fact would lead to the designation of a metric even if secondary source data or a determination by an authoritative source is not considered to contribute to the metrics. If this is true, it would be beneficial to provide a list or table of the data sources that are considered acceptable in the carcinogenicity determination. For further consideration, especially in the context of prioritization, would be how the sources of information currently noted as IG flags could contribute in a more quantitative manner to the metric as I would expect there could be substances missed through this exclusion (Table 10 indicates that only 3% of substances on the TSCA active inventory have carcinogencity data). Perhaps there is a strategy to look more closely at those substances with the IG flag that do have information other than an IARC classification or 2-year cancer bioassay, but this is not clear based on the current description. In that case, the value of 0 given in the "absence of data" may be a little misleading. Some further clarity here would be helpful.

Including further detail on how NAM will be used in the workflow would be useful. For example, in Section 5.5 and Table 11, including chemicals that use BER and TER as the domain metric and how these translate to positions on the graphical visualizations would be interesting. As a pre-prioritization exercise NAM may also be used to begin to identify those substances of potential emerging concern i.e., those

that have hazard flags, potential for exposure but for which higher tier information is not available. This could also support the identification of information gaps and accordingly, research and data needs.

In the case of the genotoxicity domain, secondary data sources were deemed acceptable and given an IG flag for awareness and presumably follow up as relevant. Why would this same approach not be considered acceptable for other hazard domains? This could be appropriate across the domains given the number of substances that will have very limited data. Perhaps this would require a metric scale of 5 to allow for these sources to have a metric = 1 rather than 0. 0 would then truly reflect no data / no flags as is done for sensitization and irritation. For consideration.

**For each of the scientific domains, there is a discussion of limitations and longer-term options. Based on your knowledge of toxicology, chemistry, and/or exposure science:**

**2C. Are the appropriate limitations and long-term options included for each domain?**

**HER Domain** - It is appreciated that the limitations regarding inhalation values are included. Could similar considerations and discussion be added for the derma routel? Exposure via the dermal route is often a key scenario and driver in risk characterizations for products. Are there complementary models or approaches that exist or that should be developed (in the future) to better include and characterize potential concern related to the dermal route of exposure?

**Carcinogenicity** -The discussion on limitations and longer-term options for the carcinogenicity domain triggered some thinking in the context on how one might better use predictive tools.
Regarding OncoLogic, some additional considerations are needed in order to incorporate the outcomes of this type of predictive system in an automated way. Some concerns would be the relevance of the flags to the parent query chemical based on OncoLogic only, as the system bases the prediction on chemical features that are common with chemical classes of carcinogenic concern and not the primary structure of interest. As such multiple chemical classes may be alerted for a single substance having many functional groups and in turn can lead to different levels of concern for the same chemical. Although it is acknowledged that OncoLogic is a valuable tool, it would be best incorporated into a consensus or weight of evidence approach with complementary predictions from other profilers.

OncoLogic Primary Classification profiler is included in the current OECD QSAR Toolbox, and can now accommodate batch runs; however, this would only provide a high-level flag as a starting point. Using the QSAR Toolbox functionalities, chemicals with OncoLogic flag would need further investigation to verify/justify the relevance of the cancer flag. This could also be done in the OECD Toolbox through the development of groups around the target chemicals but would require some expert driven evaluation.

Of note, (although likely a well know point) a lack of flag shouldn't or cannot be interpreted as absence of genotoxic or carcinogenic activity as the domain of OncoLogic is limited by its chemical classes.

Other profilers or databases in the Toolbox that might be considered in the development of a more automated approach include:

- Carcinogenic Potential Database (CPDB)
- Genotoxicity & carcinogenicity ECVAM database
- ECHA REACH database
- Cell transformation assay ISSCTA database
- Carcinogenicity & Mutagenicity (ISSCAN) database
- Carcinogenicity (genotox, nongenotox ) alerts by ISS

It is acknowledged that the suggestions above may be beyond the goal of this domain and would require some development and validation work however may provide a more substantiated approach in the longer term for data poor chemicals.

**Susceptible Populations** - Agree that a limitation of the susceptible population domain is the sole focus on children. Expanding the metric to include additional susceptible populations would be of great value. Workers is mentioned; other populations for consideration could be pregnant women, sex-related susceptibilities, geographical location / hot spots and socio-economic considerations.

The approach mentions that data collection is ongoing to expand this domain. Kudos to the EPA and ORD for their tremendous efforts related to data collection and curation. These efforts have important utility for risk assessment programs beyond the EPA.

**2D. Are there additional long-term options that could be included?**

In addition to the above, a long-term consideration could be to continue to work toward integrating broader information sources such as through the development of natural language processing approaches. This would continue to expand the data sources and enable the screening of other published literature with the goal of getting a better idea on the amount of supporting information (even if in an qualitative manner) that could be available to support assessment.

Another consideration is complementing the approach with the ability to identify groups/clusters of substances that may warrant further exploration rather than the more single substance approach outlined in the current proof-of-concept.

**3. INFORMATION AVAILABILITY**

**3A. How clearly and concisely are the descriptions of purpose and methodology of the information availability presented? Please identify areas where additional clarity is needed.**

It is clearly stated that the information availability domain is designed to automatically evaluate substances based on both the number and types of studies available. The manner by which the type of studies for each chemical is taken into account and how the amount of information available impacts the four modifying criteria could be more explicitly outlined in section 5.4. My interpretation of the description is that a value of "1" will be given to each scientific domain for which there is any single piece of experimental data to a maximum of 8 points for substances with a complete data profile in the context of the approach. This is equivalent to those chemicals for which there is an authoritative human health risk assessment. Whether "experimental data" implies in vivo experimental only is not clear. This could be defined for added transparency. An outstanding question is how each of the four modifying criteria are applied to the information availability metric derived based on the scientific domains? A low and high chemical specific example to illustrate the calculation would be informative.

**4. RESULTS and CONCLUSIONS**

**4A. Are the results of the TSCA POC clearly described and presented? If no, please identify areas where clarity is needed.**

The results of the POC clearly demonstrate the strength and consistency of this automated data driven approach when compared against chemicals with a previously characterized level of concern. Illustrating that the approach can distinguish high and low priority chemicals, when experimental data is available, using the various data plots and then providing specific chemical examples for the

calculation of the scores provides further transparency. The addition of the plot of frequency distribution of the IG flags for each of the scientific domain metrics for the POC238 set of substances was particularly interesting and gives a solid perspective on the actual data situation at hand even for what is likely a subset of the more data rich or better characterized (e.g. SCIL) chemicals. This is also acknowledged in the POC report.

Although the overall evaluation and communication of the results are well written and clearly presented, I offer a few comments for consideration.

In the opening sentence it is noted that out of the active TSCA inventory of 33, 092 substances only 15, 987 are unique organic chemical substances. If this is the defined chemical space of applicability for the PICS approach that could be stated. Also, if true, and mixtures or chemicals with greater complexity are not included, then future work may be to explore ways that an automated screening approach could be applied to the other 52% of chemicals on the active inventory. This is an area where there is likely an imperative need to apply the various NAM tools and grouping approaches that have been developed and demonstrated to have application to begin to address the more complex contexts related to screening and assessment.

Based on figure 16 (distributions of metric scores for selected substances) the whiskers span of the distributions are very large in some cases. What was interesting is that in the case of the TSCA high and the TSCA low, the lower and upper bounds for each respectively, do not overlap suggesting that it might be possible to suggest regions within the priority matrix that result in high, moderate and low priority for further work. Although this part of the evaluation is implicit in the analysis there is not a discussion on where soft thresholds could be placed on the plots to inform future activities. It may be of interest to think of these types of thresholds, or zones, to inform both priority for prioritization for risk assessment as well as priority for possible research and data generation.

The addition of the examples to illustrate the process is valuable. The example chemicals are shown to clearly separate high vs low based on the differential values of the scientific domain however based on the HER metric values of 2.7 and 2.3 for benzene and 3-methoxybutyl acetate, respectively, they are similar. In this case, the driver for the high designation is flags across hazard domains, including cancer classification, however this doesn't necessarily mean that the repeat dose toxicity flag is insignificant for 3-methoxybutyl acetate in terms of possible toxicity of concern. Could the impact and utility of the HER (or BER, TER) be further discussed in the overall context of the SDM? The chemical examples included focus on chemicals with fairly clear outcomes. Would it be possible to include another example, or 2, to also illustrate the application and impact of the BER and/or TER on outcome?

Further to the previous comment, perhaps another level of priority in the context of endpoint severity could be introduced to distinguish priority substances with existing classifications (e.g. CMR) – again just food for thought as this might be considered beyond the scope of a pre-prioritization effort.

**4B. Do the results presented adequately support the conclusions? If no, please identify and explain those issues.**

Yes. As outlined above, the selection of chemical space was relevant and appropriate to address the aim and purpose of the approach. Further, the level of analysis conducted was in-depth enough to demonstrate with examples that the conclusions are well supported for those chemicals that have *in vivo* data. The area that could be further expanded on is the inclusion of experimental data other than animal guideline studies in the case of some of the scientific domains however this does not contradict the fact the results presented support the conclusions.

## 5. EDITORIAL OR ADDITIONAL COMMENTS

**5A. Please provide any editorial or additional comments you would like to make here. These should be any comments that are not in direct response to the technical charge questions above.**

Editorial or additional comments have been included in **Table 1** below.

The existing substances bureau has also been conducting a review of data sources that are relevant for hazard identification and assessment. A supplementary excel document (**Table-2-14 Jan 2020.xlsx**), including a crosswalk of data sources in ToxValDB with those included in our in-house search strategy has been shared in case there are other data sources that may be incorporated into the EPA automated approach. In the table, those sources highlighted in red may be novel sources to consider. Some caveats are that this list has not been carefully scrutinized in detail to rule out any possible redundancies to all sources that the EPA may have already included and there are some sources that may not be amenable to automation as bulk data downloads are not possible. This would need follow up from EPA scientists.

**Table 1.**

| Section / Page / Line Number | Comment |
|---|---|
| 4.1 / pg 12 / 294 | Figure 1. There is added value in introducing the scheme at a high level as an introduction to the overall approach however the elements listed within the figure are too small to read in the schematic. Suggest increasing font size and perhaps adjusting arrangement in the vertical directions to allow for each of the SDM and IAM figures to be slightly increased in size for readability. |
| 5.0 / pg 15 / footnote 25 | Please check formatting – appears that there may be additional numerical values in the text (footnotes within footnotes?) (e.g. a direct query such as SQL20, or webservice APIs21. EPA's National Center for Computational Toxicology's Chemistry Dashboard22 is one of the several examples of a Type 1 source. The Chemistry Dashboard integrates information across various sources mapped to an expert-reviewed chemical structure23) |
| 5.0 / pg 15 / 80 | It is noted that SOPs are being implemented in a software system – Can the system be named? Is it public software or an in-house system? |
| 5.1 / pg 15 / 392 - 395 | DTXSIDs are appropriate identified however in our experience the InChI keys tend to be the most reliable substance identified. The EPA may want to consider including these as a secondary or complementary source of identifiers for mapping to CAS numbers. |
| 5.1 / pg 16 / 418 - 419 | The statement related to overall information availability implies that the exercise was conducted for the entire inventory, is this correct? If yes, could quantification of this statement be added? |
| 5.2 / pg 17 / 452 | Please define "…for making a determination…." – is this a determination of potential hazard or speaking to the overall determination of high vs low priority for each of the scientific domains? |

| | |
|---|---|
| 5.2 / pg 19 / Figure 4 | The tiered approach to data selection and application is well outlined however, I offer a few suggestions for further refinement.<br>- Exposure estimate? I understand this as estimate coming from ExpoCast only – what could lead to a "no" estimate? And how does a value of zero impact the overall outcome for the HER if that is still intended to be described based on flow in figure 4?<br>- Suggest to use a third colour box to outline IG flag<br>- Suggest that "lowest appropriate[2] POD" be defined – I suspect this was intended but it appears to be missing in the Figure description. I would also suggest that these are different for in vivo and in vitro PODs respectively. |
| 5.2 / pg 20 / 497 | For purposes of priority setting would it not be more protective from an early screening perspective to use 95$^{th}$ percentile estimates to capture possible susceptible populations as well as general population (median estimates)? |
| 5.2 / pg 20 / 500 | It is noted that "other routes of exposure are included if the units had been converted appropriately…". Dermal exposure is often a key driver for concern for products used by consumers. It is understood that dermal studies are often lacking however, is it possible to include IG flags for those chemicals where the exposure models predict likely dermal use / exposure scenarios? This could be an important trigger for further evaluation to ensure critical exposure scenarios are not missed in the screening step. |
| 5.2 / pg 20 / 506 - 508 | Additional publication for consideration to add to references - Paul Friedman K, Gagne M, Loo LH, Karamertzanis P, Netzeva T, Sobanski T, Franzosa JA, Richard AM, Lougee RR, Gissi A, Lee JJ, Angrish M, Dorne JL, Foster S, Raffaele K, Bahadori T, Gwinn MR, Lambert J, Whelan M, Rasenberg M, Barton-Maclaren T, Thomas RS. Utility of In Vitro Bioactivity as a Lower Bound Estimate of In Vivo Adverse Effect Levels and in Risk-Based Prioritization. Toxicol Sci. 2020 Jan 1;173(1):202-225. doi: 10.1093/toxsci/kfz201. PMID: 31532525; PMCID: PMC7720780. |
| 5.2 / pg 20 / 537 | Should the first term read log10(HER/BER/TER)? |
| 5.2 / pg 20 / 544-545 | If the value of zero is given as a result of no exposure information but there are (high) hazard flags is there a specific annotation that could be provided to direct the info gathering flag? I am thinking in the context of being able to better inform targeted info gathering and/or prioritizing info gathering and/or data generation efforts. |

| | |
|---|---|
| 5.2 / pg 20 / Table 1 | For clarity suggest including in the table interpretation of the continuum, i.e. 1 = highest HER (lowest concern); 4 = lowest HER (highest concern) |
| 5.2 / pg 22 / 555 - 556 | "…uses hazard information from *in vivo* repeat dose studies." It should be clear that "repeat dose" studies includes a broader scope such as reproductive and developmental studies. Perhaps a specific note to this nature earlier in the section would provide useful clarification regarding the breadth of study types included. Also, is this information restricted to guidelines studies? I suspect not necessarily if PODs that are available from an authoritative regulatory agency may be used, but this would also be a useful detail to include. |
| 5.2 / pg 22 / 568 | "…used to identify a minimum potency value showing bioactivity" – consider adding that although this is an area for refinement current evidence supports that this global bioactivity approach is protective/conservative and that further efforts to refine may provide additional pathway specific PODs (increase relevance / credible). |
| 5.2 / pg 22 / 570 | "…other sources of high-throughput bioactivity data, …". Suggest providing some example to pre-empt the application, e.g. high throughput transcriptomics, high content phenotypic data and others as relevant. |
| 5.2 / pg 23 / 585 - 587 | Carcinogencity Domain - In addition to the sources noted here (e.g., IARC, IRIS) it is suggested that GHS classification categories could also be included here as a flag indicative of the ability of an agent to cause cancer. I note that GHS is later noted as a flag for sensitization and irritation and would suggest that it could also be used for other classifications. |
| 5.2 / pg 25 / Genotoxicity Domain | Relevant citation that may be of interest:<br>Catrin Hasselgren, Ernst Ahlberg, et al. Genetic toxicology in silico protocol,<br>Regulatory Toxicology and Pharmacology, Volume 107, 2019, 104403,<br>ISSN 0273-2300,<br>https://doi.org/10.1016/j.yrtph.2019.104403.<br>(http://www.sciencedirect.com/science/article/pii/S0273230019301655) |
| 5.2 / pg 29 / Table 3 | Metric 1 and 2 – should these both indicate predicted "or" measured. Or is there an intentional difference between 1 – predicted and measured with 2 0 predicted or measured? |
| 5.2 / pg 32 / 772 | Typo - …"calculated in the same way as for human HER… |
| 5.5 / pg 50 / 1269 | Typo - benzene, 70.5~~.0~~ |

| | |
|---|---|
| 5.5 / pg 50 / 1370 | The PICS replies on a large database developed from many source databases. If there is a need for ongoing screening and prioritization in a cyclical manner is there a proposed frequency and plan for updating the data sources? Or is the workflow pulling directly from the primary sources as possible? |
| 5.5 / pg 55 / Table 11 | A few comments on Table 11.<br>- Benzene HER metric in the text indicates 2.8 and is 2.7 in the table.<br>- HER repeat dose values – it is not clear how these values are derived. Suggest adding explanation to the narrative comparison of the 2 chemicals. |

**COMMENTS SUBMITTED BY**

**Weihsueh A. Chiu, Ph.D.**

Professor, Department of Veterinary Integrative Biosciences
College of Veterinary Medicine and Biomedical Sciences
Texas A&M University
College Station, Texas

**External Peer Review of *A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA***

## 1. OVERALL QUESTIONS

**Based on your knowledge and understanding of toxicology and/or exposure, chemistry, and risk assessment, please comment on the overall TSCA POC document.**

**1A. Does this document address the purpose and aims as laid out in the introduction?**

Yes, overall, this document addresses the purpose and aims as laid out in the introduction. However, a major limitation is in how the 238 POC238 substances were selected, and the degree to which they are representative and the results generalizable. Some of the "expected" results for the POC238 are described on pages 52-53, and page 57 (paragraph lines 1352-1365). However, the fact that this is at the very end suggests that this reasoning may be somewhat post-hoc. Thus, while the POC238 may be a useful exercise in showing the feasibility of the process involved in the PICS approach, it has less utility with respect to showing that the results are useful. For instance, more formal analysis of the discriminatory power of the PICS approach – including its stated goal of allowing more false positives while reducing false negatives – may be useful before proceeding to apply it across the whole TSCA inventory. Additionally, perhaps a depiction of the chemical space covered by the POC238 as compared to the entire inventory may be useful (e.g., principal components, phys/chem descriptors such as molecular weight, log P, polar surface area, etc.).

**1B. Are the ideas presented throughout the document clear and presented in a logical manner?**

Yes, overall, the ideas in the document are clear and logically presented.

**1C. Is the method described in this document appropriate to be scalable to the thousands of chemicals on the TSCA inventory?**

Yes, overall, this method appears to be scalable to the thousands of individual chemicals on the TSC inventory. However, it should be noted that because of the QC requirements, it is not "fully automated," but will still require knowledgeable staff to implement. Additionally, it was noted that "the majority of the substances on the inventory are mixtures of varying complexity" – so it is very unclear that this approach is scalable to address those substances.

**1D. Is this approach adaptable to other large-scale chemical prioritization efforts other than for TSCA?**

Yes, overall, as a whole, this method appears to be adaptable to other large-scale chemical prioritization efforts other than TSCA, as long as only individual chemicals are prioritized.

## 2. SCIENTIFIC DOMAINS

**Based on your knowledge of toxicology, chemistry, risk assessment, and/or exposure science, please comment on the evaluation, workflow, and metrics developed for the individual scientific domains in the TSCA POC.**

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

Overall, the decisions in each domain-specific evaluation appear logical and based on sound science.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

There are several significant issues that I would like to elaborate on:

i. First, the PICS approach should strongly consider adding QSAR-derived PODs to the tiered workflow for the Human Hazard-to-Exposure Ratio domain. For instance, Li et al. (2020) [https://doi.org/10.1289/ehp6483] recently performed screening assessments that incorporated PODs from the Wignall et al. (2018) [https://doi.org/10.1289/ehp2998] QSAR model for toxicity values. Similarly, Jolliet et al. (2020) [https://doi.org/10.1111/risa.13604] also used this QSAR model in its high throughput risk and impact screening approach when in vivo values are not availab le. Additionally, it was shown in Wignall et al. (2018) that the QSAR model performed better (in terms of precision and accuracy) at predicting PODs than the ToxCast+HTTK approach employed in PICS. It should also be noted that QSAR is used in some of the other domains (e.g., genotoxicity, bioaccumulation), so there should be no reason to exclude it here.

ii. Second, the issue of metals and metalloids may need special treatment. From a screening approach (minimizing false negatives, allowing more false positives), it may be more appropriate to simply group metal and metalloid-containing compounds by the element of metal or metalloid it contains. Effects of speciation and different chemical forms is a level of detail that is probably better conducted under the "expert review and analysis" step.

iii. The determination of when data are "negative" (e.g., carcinogenicity and genotoxicity domains) needs to be made more transparent. In particular, the distinction between "inadequate"/"inconclusive" and "evidence of low likelihood" is not described. It is generally very difficult to "prove a negative" so my suggestion for both of these is that for "evidence of low likelihood" be removed as a category. Thus, "inadequate" or "inconclusive" would be the lowest level with data, and then a separate category of course for "no data." Additionally, it seems that "inadequate" or "inconclusive" would also merit some sort of flag for information availability, though not to the same degree as "no data."

iv. For the susceptible population domain, it is unclear how the "cutoffs" were determined between the value from 1-18 and the metric of 1-4. See comment below about using percentiles.

v. For persistence, the relationship between the experimental half-lives and the half-lives in the persistence criteria column is unclear. Why would biodegradation half-life of 2.75-5 weeks have the same rating as persistence half-life of >180 days? What is the justification for this correspondence?

vi. Finally, particularly for the "Ratio" metrics but maybe also for some of the others (maybe "Susceptible population, for instance), I wonder whether a percentile-based metric would be more useful, since we are really talking about relative ranking throughout the entire process. The problem with the direct use of the ratio metric is that, even on log scale, a few "outliers" can mean that the results are bunched up.

**For each of the scientific domains, there is a discussion of limitations and longer-term options. Based on your knowledge of toxicology, chemistry, and/or exposure science:**

**2C. Are the appropriate limitations and long-term options included for each domain?**

Except for those issues identified above in 2B, the appropriate limitations appear to be included for each domain.

**2D. Are there additional long-term options that could be included?**

It may be useful to include a discussion of what options are available for mixtures for each domain.

## 3. INFORMATION AVAILABILITY

**3A. How clearly and concisely are the descriptions of purpose and methodology of the information availability presented? Please identify areas where additional clarity is needed.**

The description of the information availability domain calculations is not very clear. It would be useful to have a more extensive flow-chart as to how the determinations are made for each domain and how they are added together. Additionally, it is odd that the IG flags "do not directly impact the information availability metric" – it seems that this information could be integrated. Finally, the 0 or 1 only for each domain seems overly coarse, so perhaps including the IG flags could provide a more graded score for each domain.

## 4. RESULTS and CONCLUSIONS

**4A. Are the results of the TSCA POC clearly described and presented? If no, please identify areas where clarity is needed.**

My main concern as to the description of the results is that by aggregating all the scores together, one loses the information on the individual domains. Additionally, this does not distinguish between cases of "high score" in a few domains and "moderate score" in many domains. The "rule" for moving to more detailed evaluation could be a combination of total score and "maximum score" (or "top 3" or something like that).

Additionally, a graphical visualization may be easier for communication. For instance, the ToxPI methodology (most recent published version: Marvel et al. 2018 [https://doi.org/10.1186/s12859-018-2089-2] has been used at EPA and elsewhere to help visualize multi-domain information. I would imagine actually that two ToxPIs – one for the Scientific Domains and one for the Information Availability – would be useful. You can still have an overall aggregated ToxPI Score, but having the "pie" visualization would better describe evidence for each domain. For instance, in example with Benzene and 3-Methoxybutyl acetate, a ToxPI for each chemical would very easily show the differences between the two chemicals in terms of the different metrics.

Finally, providing some idea as to the distribution across chemicals for each metric would be useful (e.g., histograms). This will help to identify whether there are certain domains where the metrics are too coarsely categorized and therefore not useful for discrimination between "higher potential" and "lower potential" substances.

**4B. Do the results presented adequately support the conclusions? If no, please identify and explain those issues.**

As I mentioned in the "overall" comments, the main issue is whether the PICS approach offers discriminatory power as to the potential for high and low priority. Thus, it seems that a more rigorous, statistically-based analysis with some prior expectation as to the results would be useful. This could be a two-stage approach, in which PICS is used on a set of chemicals, but the results blinded, and then the expert judgment process is applied to ALL those chemicals, and then the discriminatory power is analyzed (how well does PICS predict "high" and "low" priority, with the desired higher sensitivity at the expense of lower specificity).

## 5. EDITORIAL OR ADDITIONAL COMMENTS

**5A. Please provide any editorial or additional comments you would like to make here. These should be any comments that are not in direct response to the technical charge questions above.**

I do not have any additional or editorial comments.

**COMMENTS SUBMITTED BY**

**Helen M Goeden, Ph.D.**

Principal Toxicologist
Health Risk Assessment Unit
Minnesota Department of Health
Saint Paul, Minnesota

**External Peer Review of *A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA***

## 1. OVERALL QUESTIONS

**Based on your knowledge and understanding of toxicology and/or exposure, chemistry, and risk assessment, please comment on the overall TSCA POC document.**

### 1A. Does this document address the purpose and aims as laid out in the introduction?

This document describes an approach (PICS), which integrates publicly available information in such a way that can efficiently provide screening level information on potential concern within the context of how much relevant information is available. The purpose of the document is not explicitly stated. By reading through section 2 one can surmise that the purpose is to describe why the PICS approach was developed and the aims are to describe how the PICS was developed and tested using a Proof-of-Concept (POC) case study. The POC case study illustrated the ability of the PICS approach to separate lower and higher potential concern chemicals while also acknowledging the limitations of the available relevant information.

### 1B. Are the ideas presented throughout the document clear and presented in a logical manner?

The document steps through each individual domain in a logical and generally clear manner. The scope of the domains and integration are quite large but the authors should be commended on writing a concise document that is quite readable. Some of the sections could benefit from additional information, especially for readers who may not be as familiar with TSCA activities. Some of these suggestions are provided below as well as in the comments on the individual metrics.

Figure 2. Schematic of the PICS Approach in Relation to Identifying High- and Low-Priority Candidate Substances – as it is currently presented indicates that PICS is not only compiling and integrating information in a useful way but that it is the only factor going into identifying a subset of the TSCA Active Inventory for additional expert review and analysis? If this is correct – what criteria is used to identify this subset? If PICS is just one of several tools the Figure should be modified.

Section 5.1 Chemical Substance Selection, Curation and Quality Control.

Since the quality of the data was not evaluated the use of the phrase Quality Control is inappropriate and should be completely avoided throughout the document. A term such as "Data Verification" appears to be more accurate and should replace the phrase "Quality Control".

The bulleted list under Chemical Substance Selection should include mention of TSCA10 and 90. Presumably the TSCA 10 are included within the first bullet "Initial proposed set of 20 high- and 20-low-priority candidate substances" and likewise, the second bullet – "Chemical substances from the 2014 update to the TSCA Work Plan" includes the TSCA 90 list.

### 1C. Is the method described in this document appropriate to be scalable to the thousands of chemicals on the TSCA inventory?

The automated method described in the document appears to be scalable to a much larger group of chemicals. The POC list, however, was enriched in the high priority regulatory substances, and the remaining chemical substances were largely selected because of knowledge of some toxicological concern. It is not clear whether the information required to calculate a valid Scientific Domain Metric

(SDM) exists for the thousands of chemicals, which would be necessary for the PICS approach to be applied in a meaningful way.

**1D. Is this approach adaptable to other large-scale chemical prioritization efforts other than for TSCA?**

The PICS approach does appear to be adaptable for other large-scale evaluations within EPA that are focused on identifying chemicals that may be of higher concern. Again, the key factor will be availability of information to inform SDM. In addition to prioritization, the approach may also be useful in identifying common data gaps across large groups of chemicals, which could facilitate research efficiencies.

## 2. SCIENTIFIC DOMAINS

**Based on your knowledge of toxicology, chemistry, risk assessment, and/or exposure science, please comment on the evaluation, workflow, and metrics developed for the individual scientific domains in the TSCA POC.**

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

Within each domain the evaluation as described was logical and were science-based. The overall decision on the selection of the individual domains appears, at least in part, to be based on past practice. Section 5.2 explains that the domains were selected based on their importance to understanding human and ecological hazard and human exposure based on past use in TSCA prioritization activities and/or statutory language in the Frank R. Lautenberg Chemical Safety for the 21st Century Act.

In particular, it is not entirely clear why separate carcinogenicity and genotoxicity domains are identified. While it is true that cancer can occur in absence of gentoxicity that in itself is not sufficient rationale for having two of the seven individual SDM in essence assigned to the same endpoint. A single potential carcinogenicity score that incorporates genotoxicity and the available cancer data would suffice. If the goal of the selected individual domains is to identify endpoints that are of particular concern developmental/reproductive hazard should also be a separate domain. Developmental and reproductive hazards can have generational impacts. The susceptible population assessments appears to focus on exposure and does not address toxicological susceptibility.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

The workflows appear logical and the concepts are easy to follow. I did not identify significant issues but do have some suggestions or comments some of the individual domains.

**Human Hazard-to-Exposure –**

Formula 1 should be formatted in equation format to increase readability. The opposing directions of the metrics will undoubtedly cause confusion - - chemical with the lowest HER (highest concern) is set to the highest domain value of 4 and highest HER (lowest concern) is set to the lower domain value of 1. This domain would greatly benefit from having an example - the chemicals with the largest and smallest HERs are mentioned in the paragraph. Using one or both as an example to demonstrate how the equation calculations work should be beneficial. Alternatively, 3-methoxybutyl acetate and benzene, the two compounds listed in Table 11 could be used as the chemical examples throughout the document.

**5.3 Scientific Domain Metric Calculation**

Combining individual domain no data chemicals with those given a value of 1 (low concern) could result in a misleading SDM value. However, this is apparently addressed in the visual by the size of the dot – as noted in the title/description of Figure 14. This statement should also be explicitly stated in Section 5.3. as it is absolutely critical context, especially when the individual domain metrics of zero are set to 1.

**For each of the scientific domains, there is a discussion of limitations and longer-term options. Based on your knowledge of toxicology, chemistry, and/or exposure science:**

**2C. Are the appropriate limitations and long-term options included for each domain?**
**2D. Are there additional long-term options that could be included?**

I commend the authors for including a limitations and longer-term options discussion for each of the domains. I have combined my comments 2C and D below for each domain.

- ***Human Hazard-to-Exposure Ratio (HER) Domain***

  There are a variety of types of PODs, with some representing no effect levels while others may represent effect levels (e.g., NOAEL 'vs' LOAEL). This is not mentioned in the discussion but distinguishing between types of PODs is critical in providing context to the magnitude of the HER. Limitations should include identification of chemical class for which in vitro bioassays (the basis of the BER) or TTC (the basis of the TER) do not perform well.

  Use of TTC values – current TTC values have a large overlap of toxicity distributions within each category, and therefore is of limited value for separating chemicals of different toxicity potential. FDA has been working on an enhanced decision tree (EDT) for several years. An overview of the improvements were presented at the Jan 2020 Tox Forum Session: Update to the Cramer et al., Decision Tree and Thresholds of Toxicological Concern to Improve Safety Assessment and Prioritize Chemicals for Testing. The presentations were publically available for a limited time and appear to no longer be available. I do have a copy which I can make available. Alternatively, Dr Stice The three TTC classes are based on 613 compounds whereas the EDT will be based on over 1900. The large overlap between classes of different toxicity potential is vastly improved in the EDT. Based on personal communication with Dr. Stice (Szabina.stice@fda.hhs.gov), who leads this effort, there are two manuscripts under development. The first contains what, why and how and was to be submitted to Regulatory Tox and Pharm fall of 2020. The second manuscript (also to be submitted to Reg Tox and Pharm in approximately Feb of 2021) will contain the actual decision tree with over 100 example substances.

  An additional tool for estimating $NOAEL_{chronic}$ is that should be considered is the application of a $LD_{50}$-to-$NOAEL_{chronic}$ or $NOAEL_{subacute}$-to-$NOAEL_{chronic}$ extrapolation factor (Kramer et al Conversion Factors Estimating Indicative Chronic No-Observed-Adverse-Effect Levels from Short-term Toxicity Data. Reg Tox Pharm 23: 249-255, 1996). The Minnesota Department of Health has evaluated this methodology and has found that it performed better than the TTC approach. The performance of TTC vs $LD_{50}$ extrapolation was summarized in a 2017 SOT poster presentation and in project reports (an executive summary of the most recent report can be found at https://www.health.state.mn.us/communities/environment/risk/docs/guidance/dwec/execsumm2015.pdf ).

  ATSDR has also assessed the use of $LD_{50}$ by assessing LOAELs, NOAELs and MRLs from ATSDR's MRL dataset (personal communication with Siwakoti, 2016 SOT poster). The strength of association between log-LOAELs and log-$LD_{50}$s in molar units was evaluated using correlation analysis and regression. The 90$^{th}$ percentile $LD_{50}$ to $NOAEL_{chronic}$ Conversion Factor was very consistent with the 95$^{th}$ percentile CF reported in Kramer et al 1996.

The Texas Commission on Environmental Quality utilizes $LD_{50}$ extrapolation to derive NOAEL-to-$LD_{50}$ ratio-based toxicity values for chemicals with limited toxicity data (TCEQ Guidelines to Develop Toxicity Factors, Sept 2015. Accessed at: Guidelines to Develop Inhalation and Oral Cancer and Non-Cancer Toxicity Factors (texas.gov) )

Regarding the exposure parameter - it is not clear why NHANES is the sole source of biomonitoring data. Biomonitoring data is the most direct measure of human exposure and data is available from additional sources such as California (Explore Results | Biomonitoring California ) and Canada ( Human Biomonitoring of Environmental Chemicals - Canada.ca ). If these sources do not allow for automated compilation of the data that should be acknowledge in the document.

- **Carcinogenicity and Genotoxicity Domains**
  While it is true that carcinogenicity may be associated with nongenotoxic as well as genotoxic mechanisms it is not clear what the value added is by having separate domains. Many chemicals will not have cancer data - nearly half of the chemicals in the test set have no cancer data. Whereas only ~8% of the test compounds did not have genotoxicity data. Chemicals with higher cancer scores also typically had higher genotox scores. From a risk perspective chemicals with no cancer data but high genotox scores would be of concern not only for cancer but for developmental toxicity as well. Development of a carcinogenicity potential domain, that incorporates both cancer data and genotox data would provide a better metric.

  Having two of the seven SDM domains focused on carcinogenic potential seems unnecessary. In addition, other health effects of high concern such as developmental or reproductive toxicity, which can have generational consequences should be consider as a separate domain. In the absence of developmental toxicity test data the EPA TEST could be used to predict developmental toxicity along with other tools.

- **Ecological Hazard Domain**
  This domain is outside my area of expertise. I have the same comment re: formula 2 as for formula 1 – put the formula into equation format and provide example(s) to demonstrate how the domain metric value is calculated.

  Although consideration only of aquatic ecotoxicity is consistent with the GHS approach ecotoxicity for other terrestrial organisms (e.g. amphibians, birds, reptiles, etc) should be pursued as a longer-term option. The human hazard domain evaluation does not address terresterial ecotoxicity concerns.

- **Susceptible Human Population Domain**
  Although susceptible subpopulation is defined as a group of individuals within the general population who, due to either greater susceptibility or greater exposure, may be at greater risk than the general population of adverse health effects from exposure to a chemical substance or mixture this domain only focuses on the exposure aspect and does not address greater toxicological susceptibility. This focus should clearly be acknowledged in the limitations. A domain name of Susceptible Human Population Exposure Domain would be more accurate. In the absence of developmental toxicity test data the EPA TEST could evaluated as a potential predictor of developmental toxicity along with other tools.

- **Persistence and Bioaccumulation Domain**
  Glad to see that EPA is in the process of adopting new approach that includes partitioning, as well as acknowledging the current inability to adequately predict bioaccumulation of ionizable compounds. In addition to identifying this inability the relative magnitude of this problem should be conveyed (e.g., many or few substances commonly exist in the ionized state at environmental pH values).

In addition to persistence (how long a substance may remain in the environment) there should be at least two additional considerations:

1) magnitude of use. Widely used substances result in a 'constant' environmental presence and therefore are persistent in the environment, and
2) the ease or difficult for removing the substance (e.g., remediation) should also be considered.

- **Skin Sensitization and Skin/Eye Irritation Domain**
  This area is outside expertise. Multiple limitations are clearly stated.

  Longer-term options – EPA OPP has a major effort to use alternative approaches for skin sensitization and skin/eye irritation. Why are these efforts (e.g., Adopting 21st-Century Science Methodologies - Replacement Strategies | Pesticide Science and Assessing Pesticide Risks | US EPA ) not mentioned here?

## 3. INFORMATION AVAILABILITY

**3A. How clearly and concisely are the descriptions of purpose and methodology of the information availability presented? Please identify areas where additional clarity is needed.**

The description of the IAM calculation is straightforward. Within the PICS approach the information availability metric (IAM) is designed to automatically evaluate chemical substances based on the number and type of studies available to inform this analysis. Missing information is flagged but the IG flags do not directly impact the IAM and only identify specific information gaps. If the chemical substance has a human risk assessment from one of six authoritative bodies it is given a point for each of the human information availability study types. However, the existence of a human risk assessment does not indicate that information is available for each of the study types. In fact, the risk assessment from some of the authoritative bodies listed may conclude that the information available is insufficient to assess potential risk to human health. Rationale for assigning a point for each human information study type needs to be provided.

It is not clear why the IAM section does not include a Limitations and Longer-term Options discussion. Limitations in the methodology and potential longer-term options for improvement should be identified and discussed.

## 4. RESULTS and CONCLUSIONS

**4A. Are the results of the TSCA POC clearly described and presented? If no, please identify areas where clarity is needed.**

The results of the POC are clearly described and presented. The results of the POC study demonstrated that while the SDM and IAM were correlated, the PICS approach was able to segregate the recently released TSCA high- and low-priority candidate substances.

Only a small fraction of the chemical substances in the non-confidential active TSCA inventory have some in vivo mammalian data and ecotoxicological data. The data included in the PICS approach was limited to publically available data and excludes industry submitted CBI studies. The PICS approach also did not include data extracted from the literature beyond what is included in the Type 1 data sources currently being utilized. Given the dearth of publically available data and the clear association between SDM and IAM (i.e., more information tends to produce a higher value) an explanation for excluding industry submitted CBI should be provided.

**4B. Do the results presented adequately support the conclusions? If no, please identify and explain those issues.**

The results of the POC support the conclusions. Automation of data gathering and compilation to more efficiently and accurately inform chemical selection is an admirable goal. Since the POC list was enriched in the high priority regulatory substances, and the remaining chemical substances were largely selected because of knowledge of some toxicological concern it raises the question of whether sufficient data exists for PICS to achieve this goal given that there may be insufficient data to inform a SDM. This concern should be acknowledged. The PICS approach may only be useful for a subset of chemicals that have at least minimal data. If data is missing from the majority of the individual domains a reasonable estimate of an SDM is not possible.

Inclusion of 2-methoxybutyl acetate and benzene as illustrations of how the process works is very important. If possible these chemicals should be used as examples of how values are calculated using Formula 1 and 2.

[Note several of the specific scores noted in the text do not match the values in Table 11. For example, HED metric for benzene is identified as 2.8 in text but 2.7 in table.]

## 5. EDITORIAL OR ADDITIONAL COMMENTS

**5A. Please provide any editorial or additional comments you would like to make here. These should be any comments that are not in direct response to the technical charge questions above.**

I have several suggested edits to add clarity to several sections.

**Executive Summary** –
Regulatory agencies world-wide are looking to efficiently integrate information on chemical substances in order to inform priorities for decisions and data requests. This document updates the US Environmental Protection Agency's (EPA) long-term strategy described in the Working Approach for Identifying Potential Candidate Chemicals for Prioritization and presents the Public Information Curation and Synthesis (PICS) approach that integrates publicly-available hazard, exposure, persistence, and bioaccumulation information for chemical substances. The PICS approach is based on two dimensions. The first dimension, Scientific Domain Metric (SDM) synthesizes information from traditional and new approach methods (NAMs) to understand the overall degree of potential concern related to human health and the environment. The second dimension, Information Availability Metric (IAM) reflects the relative coverage of potentially relevant human health and ecological toxicity and exposure information that could inform level of effort and resources that may be needed to evaluate that specific substance. The PICS approach is not designed to replace the prioritization process described in TSCA but aims to increase efficiency and focus expert review on substances that may have a greater potential for selection as a high- or low-priority candidate.

A proof-of-concept case study was performed by applying the PICS approach to a subset of the TSCA active inventory. The results demonstrate that the approach discriminated between high- and low-priority candidate substances and identified potential information gaps. The PICS approach may be applied to large numbers of chemical substances and is an important tool for efficiently integrating and synthesizing large amounts of publicly-available information. Aspects of the approach could also be adapted and applied to other prioritization decision contexts (e.g., biosolids).

**Figure 1. Schematic of the PICS approach**
This figure should be simplified to provide an initial 'big picture' view of the approach. Strongly recommend providing a simplier and more general visual. The IAM part is completely unreadable due to the level of detail provided, a less detail visual could be used accompanied by a footnote that a more details are provided in Figure 12. Presenting the results of the POC without the accompanying details can create confusion. TSCA 10 and TSCA 90 are not even defined until page 51. With a lower level of detail in the text additional visual enhancements that convey that higher metric scores convey higher concern/greater levels of relevant information.

**Section 4.1, last paragraph needs to be broken up to enhance readability:**
The PICS approach is based on two dimensions allowing visualization and separation of the chemical substances along each dimension (Figure 1). The first dimension reflects the overall degree of potential concern related to human health and the environment and is the integration of the individual results from the domain-specific workflows. In the PICS approach, this dimension is referred to as the Scientific Domain Metric (SDM).

The second dimension reflects the relative coverage of potentially relevant human health and ecological toxicity and exposure publicly-available information that could inform level of effort and resources that may be needed to evaluate that specific substance. This dimension is referred to as the Information Availability Metric (IAM). The level of effort and resources is typically context specific and informed by expert judgment; however, an expert driven approach is not scalable to apply to the thousands of substances on the TSCA active inventory at the initial screening stage. Therefore, a set of modifying criteria were used to inform the set of potentially relevant human health and ecological toxicity information. The modifying criteria were modeled after considerations used in the EPA New Chemicals program and include a combination of functional use considerations, environmental half-life, water solubility, molecular weight, and whether the chemical substance is an exempt polymer. The existence of an authoritative human health assessment would also contribute to this metric. In the PICS approach, the summary result constitutes IAM.

The SDM and IAM are combined into a graphical representation of the PICS approach for the substances on the TSCA active inventory. In response to public comments, the PICS approach moved away from the defined 'bins' of chemical substances that had been proposed in the Working Approach. The PICS approach does not determine what a result for a specific chemical substance represents, rather it provides a synthesis of the public information available for individual substances.

**COMMENTS SUBMITTED BY**

**Kerry W. Nugent, Ph.D.**

Principal Scientist
Australian Industrial Chemicals Introduction Scheme (AICIS)
Gymea Bay, New South Wales, Australia

**External Peer Review of *A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA***

**Background**

This review takes into account experience gained during prioritization and assessment activities undertaken by the Australian industrial chemicals regulator. This organization and the relevant program changed names on 1 July 2020. The past work referenced herein was undertaken within the Inventory Multitiered Assessment and Prioritisation (IMAP) program of the National Industrial Chemicals Notification and Assessment Scheme (NICNAS). Future work is under the Evaluations program of the Australian Industrial Chemicals Introduction Scheme (AICIS). The Evaluations program is effectively a continuation of the IMAP program, with revisions of the chemical selection methodology, to account for most of the highest priority chemicals being addressed under IMAP.

## 1. OVERALL QUESTIONS

### 1A. Does this document address the purpose and aims as laid out in the introduction?

I am of the opinion that the document is very informative. It lays out a methodology for prioritizing chemicals for assessment in an automated manner, to reduce the amount of time-intensive manual data collection and review. This will only be required for chemicals selected by the automated process, for final decision making. It addresses methodologies for integrating data from animal experiments with data from new approach methodologies (NAMs), quantitative structure activity relationships (QSAR) and authoritative classification systems.

One critical aspect is that the document lays out what the approach is not intended to do. When this is taken into account, it is clear that the tools used in the PICS approach are highly relevant.

### 1B. Are the ideas presented throughout the document clear and presented in a logical manner?

The background information on the development of the ideas for prioritization of TSCA active chemicals is very useful for understanding the context of the remainder of the document. Following this, the basic architecture of the PICS approach is laid out. The individual constituents of the PICS approach are then described in detail, followed by a worked example. This reads clearly and was understandable on the initial reading. However, the Excel worksheet demonstrating the entire proof of concept working is difficult to follow and could be redesigned to maximise the logic flow of the presentation.

The approach has one novel and very useful component. That is the use of two separate dimensions, effectively one of known risk, and one of the available information. This allows decisions to be based not only on the likely risk, but also on the ability to fully characterize the risk.

### 1C. Is the method described in this document appropriate to be scalable to the thousands of chemicals on the TSCA inventory?

This may be a problem given the current day data landscape. Given that the phase-in period of the European Union (EU) REACH program is now effectively complete, the availability of more animal data for incorporation in the approach is unlikely to increase dramatically. This leads to the situation pictured on Page 53, where the information availability metric for TSCA active chemicals is very much lower than that for the proof-of-concept set. This leads to a corresponding decrease in the scientific domain metric, not because the larger group of chemicals (TSCA active) is inherently safer, but because many of the chemicals have a default position for many hazards, associated with information gathering flags.

The future developments in QSAR predictivity and NAM information collection will go some way towards addressing this issue, although there remains a hurdle relating to quantification of results from these methods, and in gaining acceptance of the relevance of these methods for regulatory rulemaking.

One aspect of the risk assessment toolkit which is not greatly used in the PICS approach is read across. This is understandable, as read across normally requires significant expert input to ensure that it is used appropriately. However, in a limited sense, read across might be automated with higher reliability. This is where substances are ionizable, and their toxicity can be assigned to a summation of the toxicity of the individual component parts. This is particularly relevant for certain endpoints such as carcinogenicity. For example, the International Association of Research on Cancer (IARC) classifies component parts, rather than individual chemicals, for carcinogenicity. An example is lead compounds. One simple way to incorporate this idea is to expand "lead compounds" to the group of individual lead compounds within the TSCA active group. More automated solutions could also be considered; I would normally consider a potassium salt to be well characterized if there is sufficient information on its equivalent sodium salt.

Specifically, for carcinogenicity, for which there is little available data, an expert system to identify chemical comprised of only low risk groups (such as alkanes, simple esters) might be of value.

**1D. Is this approach adaptable to other large-scale chemical prioritization efforts other than for TSCA?**

The split between likely risk and data availability to fully characterize the risk should be useful in other chemical prioritization exercises, for example of environmental chemicals or pesticides, as well as in other jurisdictions for industrial chemicals. However, in a broader sense, this may be useful beyond chemical prioritization, with adjustment of the scientific domain tools.

## 2. SCIENTIFIC DOMAINS

### Human Hazard-to-Exposure Ratio Domain

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

This looks appropriate. It uses very new exposure estimate developments from ExpoCast and will serve as a stringent test of this methodology. I agree with the chosen group of exposure pathways. However, I consider that it would be better to include worker exposure under this domain, rather than as a sensitive population, where there is some incongruity. Dermal toxicity and inhalation toxicity are very relevant but may require different exposure models to be fully utilized. Reproductive toxicity point of departure (POD) estimates may also be used within this domain.

Use of TER as a conservative substitute for missing POD values is supported.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

Workflow appears appropriate. The ranking system, Formula 1, seems a convenient way of assigning scores that reflect the relative risk of the chemicals. There is a minor issue with Formula 1; it would be better to include within the formula an indication the maximum and minimum POD values are across the entire set, rather than just including this in the note below. Possibly a subscript "global" might be appropriate.

**2C. Are the appropriate limitations and long-term options included for each domain?**

There is discussion of use of acute toxicity data. This often relates to different modes of toxicity to those seen in repeat dose studies and should probably be used to develop their own HER values, in conjunction with relevant short term exposure scenarios, to fully consider chemicals such as carbon monoxide.

If sufficient exposure scenarios are included (including inhalation), the difference between systemic and local effects in animal studies by inhalation is of less importance.

The value of BER values to this domain is likely to increase as high throughput bioactivity data gathering is refined and in vitro/in vivo extrapolation methodology improves.

**2D. Are there additional long-term options that could be included?**

Development of further ExpoCast models to fully utilize dermal and inhalation data, particularly for workers, and selection from this expanded range of HER values.

### Carcinogenicity Domain

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

I agree with treating carcinogenicity separately from genotoxicity, in part because of the existence of non-genotoxic modes of carcinogenicity; also to the great public concern about chemical carcinogenicity. Separation of the carcinogenicity and genotoxicity domains gives increased weighting to genotoxic carcinogens, but I consider this to be a positive outcome.

This is an extremely difficult metric, due to the severe lack of data, the difficulty in predicting non-genotoxic carcinogenicity, and the lack of potency considerations in the main available data sources (classifications). Classifications are based on the quality of available evidence, rather than carcinogenic potency. There are some issues relating to this which cause difficulty in maximally utilizing this domain.

As an example, I will use the chemical 3,3'-dichlorobenzidine, which is one of the chemicals included in the POC set. Animal experiments suggest that this chemical is probably similarly potent to the IARC Class 1 chemical, benzidine. However, there are no quality human data for this chemical, probably due to most cohorts having exposure to carcinogens apart from this chemical. Accordingly, IARC considered it to be Class 2B. There is a Globally Harmonised System (GHS) harmonized classification which has been developed within the EU, which considers it to be GHS Class 1B (more similar to IARC Class 2A). An EU harmonized classification should be considered similarly reliable to an IARC classification.

Taking this into consideration, it remains hard to envisage a better scientific basis for this determination.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

As discussed above, the simple workflow based on existing classifications seems to be the best available option. Higher level data, such as slope factors, which could serve as potency-based indicators, are available only for a prohibitively small set of chemicals.

**2C. Are the appropriate limitations and long-term options included for each domain?**

The limitations of current automated prediction systems are an important focus of ongoing work.

**2D. Are there additional long-term options that could be included?**

Two additional steps could be considered. First is expansion of group classifications by IARC (for example, lead compounds) into a set of classified chemicals within the TSCA active set. Second, there are very few reliable data available on "low likelihood of carcinogenicity", and it may be possible to develop a simple expert system to identify chemicals for which we consider the likelihood low. This would be more straightforward than full development of a prediction system.

Development of an expert system based on principles such as the Benigni-Bossa rules could also serve as a screening level prediction tool.

## Genotoxicity Domain

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

It may be of value to include additional genotoxicity test result types, rather than limit the data set to three study types. Studies such as sister chromatid exchange, unscheduled DNA synthesis and comet assays (in vitro and in vivo) would help to enlarge the data set. Inclusion of QSAR predictions, particularly those which utilize predictions of metabolism, is appropriate.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

I agree with the workflow. Chemicals classified as known mutagens should be specifically flagged. Where there are positive mutagenicity flags, it is appropriate that the PICS system should not count positive and negative results – these tests often relate to different modes of mutagenicity and in is normally not the case that uniformly positive results are obtained even for known mutagens. This determination should be left for expert consideration.

**2C. Are the appropriate limitations and long-term options included for each domain?**

The report notes than genotoxicity data are reasonably available, but often in secondary sources. It is appropriate that expert judgement is invoked by the scores given when any positive results are reported.

**2D. Are there additional long-term options that could be included?**

It is probable that data will continue to be generated for this endpoint, whether by in vitro methods or by QSAR.

## Ecological Hazard Domain

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

It is noted in the report that this domain is restricted to hazard, and data limitations restrict it to aquatic toxicity. The effect of missing data (acute and chronic, three trophic levels) is not adjusted by uncertainty factors but will be evident in the Data Availability Domain. One key consideration is that QSAR predictions of aquatic hazard data have a particularly long history and are therefore comparatively advanced.

The use of a factor of 10 for comparison of acute and chronic POD values is supported.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

The ranking system, Formula 2, seems a convenient way of assigning scores that reflect the relative hazard of the chemicals. As for Formula 1, I would like to see a subscript "global" against the maximum and minimum POD terms in Formula 2. It is pleasing to see that POD values above the maximum water solubility of the chemical are discounted by giving these chemicals a score of 1.

**2C. Are the appropriate limitations and long-term options included for each domain?**

Key limitations are described. Firstly, an automated system of environmental exposure estimation is flagged as a high priority. This would improve the quality of the environmental score by giving it a risk basis. Secondly, the limitation to aquatic data is raised. This is much more difficult to address, as there is paucity of data on other environmental compartments, and exposure estimation for these compartments is likely to be difficult. Consistent with current environmental screening, it is probable that the important aquatic compartment will remain the main focus. However, chemicals having terrestrial toxicity (animals or plants) are likely to generally have higher scores in the human toxicology sections, or aquatic toxicity, and expert review of terrestrial data may occur in higher level screening.

**2D. Are there additional long-term options that could be included?**

Environmental assessment is not my main area of work, and so I am not aware of forthcoming developments which may add to this approach.

## Susceptible Human Population Domain

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

Similarly to carcinogenicity, this is a very difficult metric, but of high public interest. The approach used here is to identify potential scenarios which may lead to higher exposure of children compared with adults. Uses of the chemical leading to their presence in exposure sources relevant to these scenarios are then identified. The approach is therefore largely a list-based expert system, which appears to be unavoidable. Accounting for different bioavailabilities for individual chemicals for the same type of exposure source would be more complex than can be addressed at this screening level.

Currently, the metric is limited to excess exposure of children, for whom the factors leading to higher exposure can be reasonably well defined. Other susceptible populations, such as pregnant women and the elderly, would be much more difficult to narrow down in this way.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

Within the limitations described above, the workflow seems appropriate. As the data do not allow quantitative estimation of the excess exposure at this screening level, the assignment of scores to represent the likely exposure differential compared with adults seems the best compromise. I do not think that disproportionate weight should be placed on monitoring data (such as dust or breast milk), as this is largely influenced by the choice of analytes in a given study. Analytes not chosen to be studied may be missed if monitoring data is too large a component of the score.

**2C. Are the appropriate limitations and long-term options included for each domain?**

As stated above, I would prefer the worker exposure to be treated in the risk based Human Hazard-to-Exposure Ratio Domain. The limitation arising from the incomplete nature of the main data sources is acknowledged, and it is probable that the quality of this domain score will improve over time.

**2D. Are there additional long-term options that could be included?**

Any identification of further factors or behaviours giving rise to greater exposure of children or of other susceptible populations should be taken into account in development of this metric. An example may be products (non-therapeutic) with disproportionate use in hospitals or aged care homes.

### Persistence and Bioaccumulation Domain

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

The methods used in this domain scoring are consistent with normal assessment practice.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

One issue which I have identified is within Table 8. I consider that a combined score of 1, which can only arise when one or the other score is 0, should be translated to a score of 0, rather than 1.

**2C. Are the appropriate limitations and long-term options included for each domain?**

There are major issues when a chemical is not a neutral organic. Some of these issues, for example for perfluorinated substances, will require more work.

**2D. Are there additional long-term options that could be included?**

It should be practicable to define relevant scoring tables for the key inorganic ions, such as Pb, Cd (suggest metric score of 4) and Na, K (probably 1). Common anions could be treated similarly, with the maximum score for cation or anion being taken forward.

### Skin Sensitization and Skin/Eye Irritation Domain

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

While these are quite different toxicological outcomes, having two separate domains would give excess weighting compared with the other toxicological domains. Compared with my previous comment about genotoxic carcinogens, in this case excess weighting is not warranted.

The scoring system is effectively aligned with the presence of GHS classifications, and requires confidence in the existing GHS classification. Alternate scoring systems would be difficult to implement, as there are multiple test methods with different scoring methodologies. These are already accounted for in the GHS classification system. However, it is valid to use raw data from more common test types (Draize, LLNA) and apply GHS rules where necessary.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

The biggest issue is with lack of reporting of GHS classifications for chemicals which do not meet the classification criteria. For this reason, lack of a classification is ambiguous, meaning either no data or not meeting criteria. This is a rationale for including the mapping of result summary sentences from REACH

dossiers. The sentences in the classification section of the dossiers ("data conclusive but not sufficient for classification") can also serve as sources for identifying negative test results. However, for the many chemicals without dossiers, this remains an issue.

I note that sensitization category 1 maps to the same result as category 1a. This is appropriate, as the older guinea pig test methods are still major data sources, and these do not allow subcategorization of positive sensitization results.

Use of the most conservative of the individual score metrics (skin sensitization, skin irritation, eye irritation) is the most appropriate way to combine the scores in this domain. However, consideration could be given to changing the sensitization score table, for example scores of 0, 1, 3 and 4, so that strong sensitizers give scores similar to the highest irritancy classifications. These chemicals can have major morbidity outcomes.

**2C. Are the appropriate limitations and long-term options included for each domain?**

Quality issues with REACH dossiers will have some impact, but the current ECHA activities will lead to improvement over time. Over-classification is not likely to be a major issue, but under-classification can lead to missed chemicals. The biggest limitation is lack of listing of negative classification results.

**2D. Are there additional long-term options that could be included?**

In the short term, the Hazardous Substance Information System (HSIS) maintained by Safework Australia can be considered a reliable source of classifications. These derive either from harmonized EU classifications or classifications determined through NICNAS assessments. Future AICIS assessment are likely to become Type 1 sources. Existing IMAP assessments are not machine readable, but the classifications available from HSIS can be used.

## 3. INFORMATION AVAILABILITY

**3A. How clearly and concisely are the descriptions of purpose and methodology of the information availability presented? Please identify areas where additional clarity is needed.**

The purpose of this domain is clear, and the decision-making process is described sufficiently. The subcategories appear logical.

One issue which does arise is that low scores may be obtained for some chemicals where assessment can still be carried out satisfactorily. These are cases where either some of the data types are not relevant (for example, carbon monoxide) or where individual hazards are so high that exposures related to other hazards will not occur (for example, potent genotoxic carcinogens). It is not clear how the second case could be addressed, but a separate category for gasses could be established.

## 4. RESULTS and CONCLUSIONS

**4A. Are the results of the TSCA POC clearly described and presented? If no, please identify areas where clarity is needed.**

The summary results in the Word document are sufficiently clear. Figure 16, in particular, is very informative. The presentation of the individual substances in the plot of Scientific Domain versus Information Availability is a visual guide to selection for assessment priority.

The use of a worked example is useful.

The Excel supporting data are not presented in a user-friendly form; identification of the key result columns is difficult, and the scores for each scientific domain could also be highlighted.

**4B. Do the results presented adequately support the conclusions? If no, please identify and explain those issues.**

The summary accurately reflects the data presented.

## 5. EDITORIAL OR ADDITIONAL COMMENTS

**5A. Please provide any editorial or additional comments you would like to make here. These should be any comments that are not in direct response to the technical charge questions above.**

I noted an interesting pair of chemicals in the test set. These are 3,3'-dichlorobenzidine and the corresponding hydrochloride. From an assessor point of view, these would largely be considered interchangeable, with possible differences only in skin and eye irritation. Given the critical importance of the carcinogenicity potential of these chemicals, these differences would be of little importance. Based on the interchangeability of data between these two chemicals, I would like to see in the longer term that the system can converge them to the same results for both scientific domain and for data availability.

**COMMENTS SUBMITTED BY**

**Edward J. Perkins, Ph.D.**

U.S. Army Senior Research Scientist
Environmental Networks and Genetic Toxicology
U.S. Army Engineer Research and Development Center (ERDC)
U.S. Army Corps Engineers (USACE)
Vicksburg, Mississippi

**External Peer Review of *A Proof-of-Concept Case Study Integrating Publicly Available Information to Screen Candidates for Chemical Prioritization under TSCA***

**1. OVERALL QUESTIONS**
**Based on your knowledge and understanding of toxicology and/or exposure, chemistry, and risk assessment, please comment on the overall TSCA POC document.**

**1A. Does this document address the purpose and aims as laid out in the introduction?**

In section 2, the introduction lays out the purpose of the document to be to describe the Public Information Curation and Synthesis (PICS) approach for integrating publicly-available information on the more than 33,000 chemical substances on the non-confidential TSCA active inventory to efficiently select substances for expert review prior to prioritization. The document aims to do this using a proof-of-concept case study to show how the PICS approach can help streamline the evaluation of chemical substances by transparently and reproducibly synthesizing available information and identify potential data gaps. Additionally, other sections (1) indicate that the document aims to demonstrate that the PICS approach can discriminate between high- and low-priority candidate substances, (4.1) aims to present potential options for future work to improve the approach, as well as caveats and limitations and (4.2) aims to describe what the PICS approach is and is not intended to accomplish.

Overall, the document does a thorough job addressing the purpose and aims described above. It would be helpful if the purpose and aims were also more explicitly and succinctly described in the executive summary rather than summarizing the document as if it were a research project. The document presents a potentially complex approach where one could get lost in the details. However, providing practical examples through the use of case studies and specific chemical comparisons was very helpful in explaining and demonstrating how the PICS approach works and can help streamline chemical evaluations.

**1B. Are the ideas presented throughout the document clear and presented in a logical manner?**

Yes, in general I thought they were presented clearly and logically. Section 4 was much appreciated as it helped to clearly set expectations for the rest of the document. The figures provide nice summaries of the different sections, but would be even more helpful if the fonts were larger (e.g. figure 1). While acronyms can be helpful to reduce overuse of words, the document could have more clarity with less use of acronyms and spelling out infrequently used acronyms/jargon. The inclusion of workflow diagrams in several of the scientific and information domain sections were informative. It would help with clarity if a high-level workflow chart figure were provided at the beginning of section 5 that illustrates the various overall steps of the proof-of-concept case - similar to figure 1, but with more details.

**1C. Is the method described in this document appropriate to be scalable to the thousands of chemicals on the TSCA inventory?**

Yes, this approach should be scalable to assessing information associated with thousands of chemicals. The demonstration examined 238 chemicals and the number of chemicals that approach can examine seems to only be limited by the ability to access the appropriate information in the structure used by

PICS (e. g. type 1 data). These limitations can be overcome by transforming data/information into the correct structures [1] prior to analysis in addition to the cleaning and curation of data.

**1D. Is this approach adaptable to other large-scale chemical prioritization efforts other than for TSCA?**

Yes, the approach seems to be very flexible. Since PICS reproducibly synthesizes available information and identify potential data gaps it could be used in many different efforts. Designation of what characteristic (e.g. good, ok, bad, very bad) 1 to 4 scale could be assigned different metrics that may be more relevant to other efforts. For example, the approach could be used to inform prioritization of chemicals of emerging concern for human health in drinking water and ecological health in the environment. The approach appears flexible enough so that other scientific domains could be added or removed where/when appropriate. For example, the addition of a scientific domain describing partitioning of a chemical into biosolids and/or soil matrix would help identify what chemicals introduced through use of biosolids may pose the greatest a health hazard or identify chemicals that should be studied more because of lack of information and potential for persistence and greater exposure.

**2. SCIENTIFIC DOMAINS**

**Based on your knowledge of toxicology, chemistry, risk assessment, and/or exposure science, please comment on the evaluation, workflow, and metrics developed for the individual scientific domains in the TSCA POC. Please explain and justify your rationale for your responses to the charge questions.**

**2A. Were the decisions in each of the domain-specific evaluations logical and based on sound science?**

In general, they were logical, well described and based on sound science. In many cases they have taken common practices and used them as different tiers of workflow.

The human hazard relative to exposure domain follows the accepted paradigm for relevance of toxicity data for human risk: in vivo human>in vivo animal repeat dose toxicity>in vitro assay>in silico (threshold of toxicological concern or TTC).

The carcinogenicity domain takes a logical approach that minimizes confusion on information by scoring if authoritative source has determined that a chemical causes human carcinogenicity, then if there is evidence of animal carcinogenicity. This approach simplifies assessment of carcinogenicity and flags chemicals that display the potential for causing cancer for expert assessment.

The Genotoxicity domain takes the logical approach of using the standard tests for genotoxicity recommended by the Organization for Economic Cooperation and Development (OECD) for in vitro experimental determination of genotoxicity as the first tier of the workflow followed by in silico prediction of genotoxicity if no test data is available. This creates a tiered assessment of measured evidence for genotoxicity> prediction of genotoxicity> inconclusive evidence of genotoxicity> low likelihood of genotoxicity> no data. Each tier of assessment is backed by sound science.

The Ecological hazard domain scheme is logical and follows a standard approach in ecotoxicological testing by first considering chemical effects on three major trophic groups (fish, invertebrate (crustacean), and a plant/algae). If no in vivo acute or chronic data is available, then Quantitative

[1] **Clarification:** Yes I meant the correct data structure. The correct structure means that the data would be already be formatted and compiled in manner such that no data cleaning or reformatting would be needed (e.g. type one data). The data would be immediately usable for analysis without further manipulation.

Structure Activity Relationships (QSAR) models are used to predict toxicity. Although, as discussed in the document, exposure is not considered in the evaluation which could lead to higher numbers of chemicals with higher hazard scores.

The Susceptible populations domain presents a logical approach for assessing if a chemical has the potential for a differential exposure between children and the general population. Media/environments where children are likely to experience a high exposure to a chemical, if it were present (e.g. breastmilk), are given higher scores than media/environments where children not expected to have high exposures (e.g. far field sources). The approach taken to sum scores for each media/environments that a chemicals may be present in is consistent with scientific understanding that exposures have cumulative effects.

Persistence in the persistence and bioaccumulation domain is based on a conservative "ultimate biodegradation" or mineralization to CO2 and water. This is a reasonable approach as incomplete mineralization could result in reduction of a chemical of concern but increase levels of daughter products that could be hazardous. Basically, ultimate biodegradation simplifies assessment by not having to consider potential hazards of daughter products. The bioaccumulation component takes a logical approach by first assessing whether or not a high-quality Bioaccumulation factor (BAF) or bioconcentration factor (BCF) is present. If no high-quality data are present then a BCF is predicted using accepted modeling approaches. A major reasonable assumption is that chemicals are neutral at environmental PHs – necessary as the modeling components can't accurately predict behavior of ionized chemicals.

The skin sensitization and skin/eye irritation takes the logical approach of tiering authoritative values at the first level, then screening level values at the second level, then QSAR or other modeling values at the third level. At each level the most conservative value is selected to ensure potential hazards are not underreported.

**2B. Do you see any significant issues with any of the tiered workflows and metrics developed for these scientific domains? If so, please identify and explain those issues.**

In general, the tiered workflows and metrics are reasonable given the purpose of the effort and limitations described in each domain area.

The workflow for the persistence sub-domain seems to be lacking in analysis of partitioning in different media, desorption rates and other factors that could increase persistence or decrease bioavailability of a chemical. It appears that Ionizable chemicals such as the PFASs would lack information in several domains that use QSARs. Specifically, PFASs would not be modeled well in domains using QSARs, especially in the persistence and bioaccumulation domain. Both of these issue/limitations were noted in the document.

As noted in the document the ecological hazard domain assessment does not include an exposure factor. This could make it more difficult to discern between hazardous chemicals. Without exposure factors, some chemicals will appear to be more hazardous than when a risk-based approach including exposure is used.

Currently the susceptible population domain only assesses exposure of children. This ignores other susceptible populations such as elderly, low income or ill subpopulations. As a result this may result in chemicals that these subpopulations are disproportionately exposed too being overlooked.

**<u>For each of the scientific domains</u>, there is a discussion of limitations and longer-term options. Based on your knowledge of toxicology, chemistry, and/or exposure science:**

**2C. Are the appropriate limitations and long-term options included for each domain?**

Yes, appropriate limitations and long-term options are included for each domain. In general, a good job is done in describing current limitations and potential longer term options for improving the effort.

Human hazard relative to exposure domain: In the description of human hazard relative to exposure depicted in figure 4 (line 480) New approach methodologies (NAM) are noted in the figure but could be pointed out in the text. Identification of the TTC as an in silico NAM would help illustrate how this class of information could be integrated into the overall scheme. Given the importance of NAMs it would be good to mention whether or not the workflow would handle other NAMs such as alternative species or *in chemico* assays. How will other models and NAMs be used in the long term?

**2D. Are there additional long-term options that could be included?**

Other long-term options for human and ecological domains would be to include assessment of specific pathways ( estrogenic, etc) or Adverse Outcome Pathways based on in vivo, in vitro and or in silico data. Development of models predictive of hazards based on structure that can provide more accurate assessments than current models. Development of methods that can predict biomagnification.

**3. INFORMATION AVAILABILITY**

**3A. How clearly and concisely are the descriptions of purpose and methodology of the information availability presented? Please identify areas where additional clarity is needed.**

In general, I feel that, given the goals of the effort, the descriptions of purpose and methodology of the information availability were very clear and understandable. The examination/comparison of the two chemicals benzene and 3-methoxybutyl acetate were very helpful in understanding the process.

On line 1269, its unclear if the number after benzene represents the SDM values- is 70.5.0 a typo or is it another item? Based on table 11 it's a typo. It may help to present SDM values the same way as the IAM values (Benzene = 93%).

**4. RESULTS and CONCLUSIONS**

**4A. Are the results of the TSCA POC clearly described and presented? If no, please identify areas where clarity is needed.**

Yes, I felt that the results of the TSCA POC were clearly described and presented. I appreciated the comparison with other chemical groupings which had different levels of information (e.g. Safer Choice Ingredients List) or had expert driven assessments ( e.g. TSCA Hi and Low). While the PICS approach has limitations and potential biases to accessible datasets and conservative estimates, PICS successfully identifies data relevant to chemical hazard and risk assessment in several area and manages to collapse this into two interpretable metrics that can be easily explored.

**4B. Do the results presented adequately support the conclusions? If no, please identify and explain those issues.**

The results of the POC do support the conclusions that the PICS approach can help inform chemical prioritization and identify possible data needs. Comparison of the POC238 to other chemical groupings ( e.g. Figures 14 and 16. TSCA 90, TSCA Hi/Low, SCIL, food ingredients) demonstrated that PICS can identify chemicals with evidence of high impact versus those with low impact. These figures also show that PICS can identify chemicals with evidence of high impact that have low information availability.

Chemicals with high SDM/ low IAM could be prioritized in future investments so that evidence can be developed to better define the hazard/risk of chemicals. Chemicals with high levels of information and evidence for high impact/hazard/risk could be priority for expert assessment due to the availability of information and need to understand hazard/risks of chemicals with high SDMS.

## 5. EDITORIAL OR ADDITIONAL COMMENTS

**5A. Please provide any editorial or additional comments you would like to make here. These should be any comments that are not in direct response to the technical charge questions above.**

Overall, I found the document clear and easy to read. EPA did a great job assembling the document and performing the POC study. I especially appreciated the flow charts describing what is being done in each section. Several acronyms have not been defined before use. For example: Need to define LLNA (Line 1068). OCSPP (line 1287).